Name:_____

CSE 544, Fall 2006, Take-Home Examination
Due: 13 November 2006, in class

Rules:

- You can use any documentation that you want to complete this exam.

- You are allowed to search for information on the Web.

- Please write clearly.

- **You are not allowed to talk to anyone about the exam.**

Question	Max	Grade
1	10	
2	15	
3	25	
4	10	
5	20	
6	12	
7	8	
Total	100	

Name:_____

1. (**10** points)   **Data Independence**

    (a) (**2** points)   What is physical data independence?

    (b) (**1** points)   What properties of the relational model facilitate physical data independence?

    (c) (**2** points)   What is logical data independence?

(d) (**1** points)   How can one provide a high level of logical data independence with the relational model?

(e) (**4** points)   Name four concepts introduced by E. F. Codd in his paper "A Relational Model of Data For Large Shared Data Banks" that are part of the relational model as we know it today.

2. (**15** points) **Schema normalization**

Your friend Bob is designing a database for a scientist. The database will serve to track tagged animals over time. Sensors placed at fixed locations will automatically record animal sightings. Bob is considering using the following schema:

```
Animals (
           tag_id :  integer,       Unique ID associated with the animal's tag
           type:  string,          The type of animal: cow, crow, turtle, etc.
           description:  string,   Describes any special characteristics of this type of animal.
           markings:  string,      Describes any special markings of this animal.
           sighting_date:  date,   The date when the animal was detected by the sensor.
           sensor_id:  integer,    Sensor ID.
           latitude:  integer,     Location of the sensor.
           longitude:  integer     Location of the sensor.
)
```

where (`tag_id, sighting_date, sensor_id`) is the primary key of `Animals`.

(a) (**5** points)   List and describe three specific examples of problems (anomalies) that might arise when inserting into, deleting from, or updating this database.

(b) (**5** points)   What is the difference between the following two normal forms: 3NF and BCNF? [Please continue on next page]

Name:_____

(c) (**5** points) Propose a decomposition of Bob's schema into multiple tables that avoids the problems you listed above (show your schema below).

3. (**25** points)   **Query optimization**

   Given the following SQL query:

   ```
   Student (sid, name, age, address)
   Book(bid, title, author)
   Checkout(sid, bid, date)

   SELECT S.name
   FROM   Student S, Book B, Checkout C
   WHERE  S.sid = C.sid
   AND    B.bid = C.bid
   AND    B.author = 'Olden Fames'
   AND    S.age > 12
   AND    S.age < 20
   ```

   And assuming:

   - There are $10,000$ `Student` records stored on $1,000$ pages.
   - There are $50,000$ `Book` records stored on $5,000$ pages.
   - There are $300,000$ `Checkout` records stored on $15,000$ pages.
   - There are 500 different authors.
   - Student ages range from 7 to 24.

   (a) (**5** points)   Show a physical query plan for this query, assuming there are no indexes and data is not sorted on any attribute.

(b) (**5** points)   Compute the cost of this query plan and the cardinality of the result.

(c) (**5** points)   Suggest two indexes and an alternate query plan for this query.

Name:_____

(d) (**5** points)   Compute the cost of your new plan.

(e) (**5** points)   Explain the steps that the System R query optimizer would take to optimize this query.

4. (**10** points)   **Transactions: concurrency control**

Consider a database with objects X and Y and assume that there are two transactions T_1 and T_2. T_1 first reads X and Y and then writes X and Y. T_2 reads and writes X then reads and writes Y.

(a) (**3** points)   Give an example schedule that is not serializable. Explain why your schedule is not serializable.

(b) (**3** points)   Show that strict 2PL disallows this schedule.

Name:_____

(c) (**4** points)   What are the differences between the four levels of isolation?

Name:_____

5. (**20** points)   **Transactions: recovery**

   (a) (**2** points)   What are STEAL and NO-STEAL policies?

   (b) (**2** points)   What are FORCE and NO-FORCE policies?

   (c) (**6** points)   In ARIES, what is the goal of the analysis phase? Please describe what happens during that phase.

(d) (**5** points)   In ARIES, what is the goal of the redo phase? Please describe what happens during that phase.

(e) (**5** points)   In ARIES, what is the goal of the undo phase? Please describe what happens during that phase.

6. (**12** points)   **Distribution**

   Please use the description of 2PC from Mohan et. al. to answer the questions below.

   (a) (**6** points)   In the two-phase commit protocol, what happens if the coordinator sends PREPARE messages and crashes before receiving any votes?

      i. (**2** points)   What is the sequence of operations at the coordinator after it recovers.

      ii. (**2** points)   What is the sequence of operations at a subordinate that received the message and replied to it before the coordinator crashed.

      iii. (**2** points)   What is the sequence of operations at a subordinate that did not receive the message before the coordinator crashed.

Name:_____

(b) (**6** points) In the two-phase commit protocol, why do subordinates need to force-write a prepared log record before sending a YES VOTE? To answer this question, use an example failure scenario. Show what happens if a subordinate does NOT force-write the prepared log record, then show what happens if the subordinate does force-write the prepared log record.

7. (**8** points)   **Replication**

    (a) (**3** points)   In eager master replication, when the master fails, why does a group of replicas need to have the majority of nodes in order to elect a primary and continue processing requests?

    (b) (**5** points)   What are the differences between eager and lazy replication? Please list differences in the approaches and differences in the properties that result. Discuss master vs group replication if appropriate.